Genetic Epidemiology 30: 48–61 (2006)

Localization of Breast Cancer Susceptibility Loci by Genome-Wide SNP Linkage Disequilibrium Mapping

Nathan A. Ellis,[1] Tomas Kirchhoff,[1] Nandita Mitra,[2] Tian-Zhang Ye,[1] Shaokun Chuai,[2] Helen Huang,[1] Khedoudja Nafa,[1] Larry Norton,[1] Susan Neuhausen,[3] Derek Gordon,[4] Jeffery P. Struewing,[5] Steven Narod,[6] and Kenneth Offit[1]*

[1]*Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, New York*
[2]*Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, New York*
[3]*Department of Medicine, University of California at Irvine, Irvine, California*
[4]*Laboratory of Statistical Genetics, Division of Epidemiology, Rockefeller University, New York, New York*
[5]*Laboratory of Population Genetics, National Cancer Institute, Bethesda, Maryland*
[6]*Department of Public Health Sciences, University of Toronto, Toronto, Ontario, Canada*

We studied the feasibility of a novel approach to localize breast cancer susceptibility genes, using a low-density genome-wide panel of single-nucleotide polymorphisms and taking advantage of large regions of linkage disequilibrium (LD) flanking Jewish disease genes in high-risk cases. With Affymetrix GeneChip arrays, we genotyped 8,576 polymorphisms in three sets of Ashkenazi Jewish breast cancer cases: a ''validation'' set of 27 breast cancer cases, all of whom carried the *BRCA2*6174delT* founder mutation; a ''field'' set of 19 breast cancer cases from male breast cancer kindreds, which simulated conditions for finding new genes; and a ''test'' set of 57 probands from breast cancer kindreds (4 or more cases/kindred), in which mutations in *BRCA1* and *BRCA2* had been excluded. To identify associations, we compared the frequency of genotypes and haplotypes in cases vs. controls by the Fisher's exact test and a maximum likelihood ratio test. In the ''validation'' set, we demonstrated the presence of a region of linkage disequilibrium on *BRCA2*6174delT* chromosomes that spanned over 5 million bases. In the ''field'' set, we showed that this large region of linkage disequilibrium flanking *BRCA2* was detectable despite the presence of heterogeneity in the sample set. Finally, in the ''test'' set, at least three regions of interest emerged that could contain novel breast cancer genes, one of which had been identified previously by linkage analysis. While these results demonstrate the feasibility of genome-wide association strategies, further application of this approach will critically depend on optimizing the density and distribution of SNPs and the size and type of study design. *Genet. Epidemiol.* 30:48–61, 2006. © 2005 Wiley-Liss, Inc.

Key words: single-nucleotide polymorphism; linkage disequilibrium mapping; breast cancer

Abbreviations used: EM, expectation-maximization; LD, linkage disequilibrium; Mbs, million bases; SNP, single-nucleotide polymorphism.

Grant sponsor: Academic Medicine Development Company; Grant sponsor: Werner and Elaine Dannheisser Fund, Lymphoma Foundation; Grant sponsor: Frankel Fellowship; Grant sponsor: Danziger Foundation; Grant sponsor: Byrne Foundation; Grant sponsor: Sloan-Kettering Institute; Grant sponsor: National Institutes of Health; Grant numbers: RO1-HL56778, CA 103394, KO1-HG-00055, and MH44292.

N.A.E. and T.K. contributed equally to this work.

Present address of Nathan A. Ellis: Gastroenterology Section, University of Chicago, 5841 S. Maryland Ave., MC4076, Chicago, IL 60637
*Correspondence to: Kenneth Offit, M.D., M.P.H., Clinical Genetics Service, Memorial Sloan-Kettering Cancer Center, 1275 York Ave., New York, NY 10021. E-mail: offitk@mskcc.org

Received 7 January 2005; Accepted 28 July 2005
Published online 3 October 2005 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/gepi.20101

INTRODUCTION

Five percent to 15% of incident breast cancer cases are thought to result from autosomal-dominant cancer susceptibility genes [Colditz et al., 1993; Slattery and Kerber, 1993; Claus et al., 1996]. However, only approximately half of hereditary breast cancer cases are the result of known disease-causing mutations in *BRCA1*, *BRCA2*, or other identified cancer susceptibility

genes [Couch and Weber, 2002]. For example, in a set of families studied by the Breast Cancer Linkage Consortium, only 41% of 83 families with 4 or 5 cases of female breast cancer demonstrated linkage to either *BRCA1* or *BRCA2* [Ford et al., 1998]. However, attempts to utilize linkage to localize other genes associated with an inherited predisposition to cancer have been hampered by genetic heterogeneity, decreased penetrance, and chance clustering [Thompson et al., 2002]. Illustrating these problems, loci identified by linkage on chromosomes 8 and 13 [Kerangueven et al., 1995; Seitz et al., 1997; Kainu et al., 2000] were not confirmed in analyses in other studies [Rahman et al., 2000; Thompson et al., 2002].

An alternative to linkage analysis for localization of cancer susceptibility genes is genetic association [Botstein and Risch, 2003; Schaid, 2004]. Because power for both linkage and association methods is weakened by genetic heterogeneity, we investigated the study of the Ashkenazi Jewish population, which is a genetically isolated population that was established from a relatively small number of founders. In a genetically isolated population, genetic drift will have reduced the number of disease-causing mutations segregating in the population; consequently, each founder mutation is present in linkage disequilibrium (LD) with a segment of the surrounding chromosome that contains a particular linear combination of alleles at the flanking loci [Shifman and Darvasi, 2001], referred to as a founder haplotype. Thus, a plausible strategy to localize new disease genes is to perform a case-control study in a genetically isolated population, testing for regions of the genome in which one specific founder haplotype is significantly more frequent in cancer cases than in controls.

For the identification of novel cancer genes, the Ashkenazi Jewish population is ideal. The limited number of founder individuals who established the population, population identity, and the large size of the present-day population combine to make possible the study of large numbers of persons with specific cancers. In addition, the size of the founder haplotypes surrounding the disease-causing mutations in the Ashkenazim (e.g., in Jewish hereditary nonpolyposis colorectal cancer, HNPCC) is often large, on the order of 1–10 million base pairs (Mbs) [Mitra et al., 2004]. In hereditary breast cancer syndrome, two Ashkenazi founder mutations *BRCA1∗187delAG* and *BRCA2∗6174delT* are each present in approximately 1% of Ashkenazi Jews, but they are present at increased frequencies in early-onset Ashkenazi Jewish breast cancer cases and in Ashkenazi Jewish breast cancer families [Neuhausen et al., 1996a; Offit et al., 1996]. Testing for the three founder mutations (including *BRCA1∗5183insC*) captures >95% of carriers with *BRCA1* or *BRCA2* mutations in the Ashkenazi population [Frank et al., 2002; Kauff et al., 2002; Phelan et al., 2002]. In persons who carry these ancient mutations, a region of 1–3 Mbs that flank each of these mutations is identical by descent from a founder individual [Neuhausen et al., 1996b, 1998]. Because the size of the founder haplotypes flanking disease genes is large in the Ashkenazi population, we reasoned that we could use a panel of single-nucleotide polymorphisms (SNPs), distributed evenly but at a low density across the genome (on the order of 1 SNP per 0.4 Mb), to identify these founder haplotypes in breast cancer cases.

An enhancement in our study design involves the use of ''enriched'' breast cancer cases selected from families that contain multiple affected persons. The power of association studies can be significantly increased by the selection of cases from families with a history of breast cancer [Houlston and Peto, 2003]. For example, assuming two controls for each case, a dominant allele with a relative risk of 2, and an allele frequency in the population under study of 2%, 1,200 unselected cases would be needed to detect the association with 95% power at an alpha of 0.01, whereas only 350 cases would be needed if the cases had been selected from families with two first-degree relatives affected with breast cancer [Houlston and Peto, 2004]. An example of this selection effect is provided by recent analyses of the CHK2 mutation 1100delC in breast cancer. This mutation is only slightly increased in frequency in unselected breast cancer cases in comparison to healthy population controls, but it is greatly increased among familial cases not carrying a *BRCA1* or *BRCA2* mutation [Meijers-Heijboer et al., 2002]. A *CHK2* mutation specific to the Ashkenazi Jewish population was recently described [Shaag et al., 2005], and we investigated whether or not an association could be detected in our case series between the region that contains *CHK2* and breast cancer.

To test the feasibility of genome-wide SNP LD mapping using a low-density SNP panel, we studied the pattern of LD in a set of 27 breast cancer cases that carried the *BRCA2∗6174delT*

mutation, and then tested the utility of the method in a clinical ascertainment of 19 cases enriched for *BRCA2* mutation carriers. Based on these results, we applied an association method to a set of 57 cases from kindreds containing four or more breast cancers but lacking a mutation in *BRCA1* or *BRCA2*. While demonstrating significant limitations due to statistical power resulting from low SNP density and sample size, these results suggest that genome-wide SNP LD mapping in enriched cases from a genetic founder population is a promising approach to the identification of novel breast cancer susceptibility genes.

MATERIALS AND METHODS

HUMAN SUBJECTS

All probands were drawn from kindreds ascertained over a 9-year period (1994–2004) in the Clinical Genetics Service at Memorial Sloan Kettering Cancer Center (New York, NY). Three ascertainment sets were assembled for this study (Table I). The first consisted of 27 Jewish females with breast cancer who were carriers of the *BRCA2∗6174delT* mutation (the ''validation'' set). This sample set was tested to ensure that the SNP set in use was sufficiently informative, and to gain insight into the strength and extent of linkage disequilibrium around *BRCA2*. To simulate gene discovery conditions that would operate if we were blinded as to genotype, we next selected Jewish probands (n = 19) from all available kindreds with two or more breast cancers, one of which occurred in a male breast cancer patient (the ''field'' set). This ascertainment was chosen because it is known that breast cancer families sampled through families with male breast cancer cases are enriched for mutations in *BRCA2* [Stratton et al., 1994; Wooster et al., 1996]. To test the application of the association method to find new genes, we ascertained a third group of Jewish probands (n = 45) from kindreds with four or more breast cancers in first- and second-degree relatives (the ''test'' set). To increase the power of the ''test'' set, an additional 12 probands who came from families that met the ''test'' set criteria were ascertained through the University of Toronto for a total of 57 cases. The cases in the ''test'' set had either undergone sequencing of the entire open reading frame of *BRCA1* and *BRCA2* (n = 19), or were tested for the Ashkenazi Jewish founder mutations only (n = 38). Because testing for the Ashkenazi founder mutations captures >95% of carriers with *BRCA* mutations in this population [Frank et al., 2002; Kauff et al., 2002; Phelan et al., 2002], undetected mutations in this set of cases should be infrequent. The New York cases were further tested for possible deletions in *BRCA1* and *BRCA2* by the semiquantitative multiplex polyermase chain reaction method [Charbonnier et al., 2000], and none were detected. All cases provided written informed consent for the use of their DNA for research into the genetic causes of breast cancer. For the control group, 23 DNA samples from Ashkenazi persons were purchased from the National Laboratory for the Genetics of Israeli Populations at Tel-Aviv University (Tel-Aviv, Israel), and another 40 DNA samples from healthy Ashkenazi Jews were chosen at random from a group of over 2,000 cancer-free persons aged 18–60 years, ascertained through the New York Cancer Project [Gruber et al., 2002]. These 63 controls were employed for all three comparisons with breast cancer case groups. The median ages and age ranges of cases and controls used in this study are shown in Table I, along with the mean numbers of

TABLE I. Mean age, age range, and number of affected relatives in subject population

				Mean (median) number of relatives with breast cancer		
Cases	N	Mean age (median)	Age range	Female	Male	Total
Validation set[a]	27	51.4 (48)	36–90	1.8 (2)	0.3 (0)	2.1 (2)
Field set[b]	19	51.5 (50)	38–68	2.3 (2)	0.7 (1)	2.9 (3)
Test set[c]	57	49.2 (48)	31–71	3.8 (4)	0.0 (0)	3.8 (4)
Controls						
Healthy	63	49.4 (49)	30–82		Not known	

[a]Fifteen cases were affected carriers, and 12 cases were unaffected carriers. For age calculation, age of diagnosis was used for affected carriers, and age at which specimen was taken was used for unaffected carriers. Family history information was unavailable for five patients.
[b]''Field'' set contained five cases that were included in ''validation'' set.
[c]Information on age of diagnosis and on number of affected relatives was unavailable for four patients.

TABLE II. Mean and median distances in kilobases between SNPs

Chromosome	Mean	Median
1	369	188
2	324	157
3	314	155
4	321	183
5	296	151
6	274	146
7	349	173
8	319	144
9	308	1,378
10	280	143
11	272	117
12	328	140
13	252	134
14	270	140
15	325	175
16	427	175
17	533	271
18	314	145
19	723	328
20	371	189
21	226	117
22	552	276
X	1,078	625
Total	384	192

female and male relatives with breast cancer in the cases.

GENOTYPING

DNA samples were prepared from blood as previously described [Peterlongo et al., 2003]. Genotyping was carried out using Affymetrix GeneChip 10K Human Mapping Arrays [Kennedy et al., 2003; Matsuzaki et al., 2004]. Briefly, the method consisted of a one-primer amplification assay performed on genomic DNA in which sequence complexity had been reduced by restriction enzyme digestion with *Xba*I. Allele-specific hybridization of the amplified probe was then performed on oligonucleotides on the array. Because two different chip arrays were used in these experiments (the early-access 10K array and the 10K array), the loci included in the statistical analysis were constituted from the intersection of the two sets of loci successfully genotyped (8,576 total SNP, not including 128 SNPs on the X chromosome). The overall genotyping failure rate in the 161 DNA samples genotyped by this method was 8.6%. The genotyping error rate was estimated to be less than 0.02% (no genotype discrepancies in 5,166 tests at 63 duplicate loci included on the genotyping chips). The mean and median distances between the loci genotyped in these experiments were 384 kb and 192 kb, respectively (Table II). The mean heterozygosity for the 161 Jewish persons tested was 33% (median, 27%).

STATISTICAL ANALYSIS

Two-sided Fisher's exact tests were used to compare the single SNP genotype frequencies between cases and controls. For haplotype analyses, we ordered all loci by chromosome position, from the telomere of the p arm to the telomere of the q arm for each chromosome in turn, and applied a sliding window consisting of n loci, where n was a number from 2–12 that we moved down the chromosome one locus at a time [Fallin et al., 2001]. In the estimation of haplotype frequencies, we included samples in which genotype data were missing according to the following scheme: for $n = 3$ or 4, we excluded samples with missing data at two or more loci; for $n = 5$ or 6, we excluded samples with missing data at three or more loci; for $n = 7$ or 8, we excluded samples with missing data at four or more loci; for $n > 8$, we excluded samples with missing data at five or more loci. Using multiple imputation for any remaining missing data, we assigned the most likely haplotypes to each individual, using the expectation-maximization (EM) algorithm [Excoffier and Slatkin, 1995]. From there, we compared estimated haplotype frequencies between cases and controls as described below. Because we selected cases from a genetically isolated population, we assumed that in some or all cases, a single founder haplotype is present that is associated with the disease. Consequently, for each group of adjacent loci, we identified the haplotype that obtained the minimum *P*-value.

To calculate *P*-values, a maximum likelihood ratio test was performed as described by the SAS Institute, Inc. [2002]. For a given group of adjacent loci, the chi-square statistic of each haplotype was calculated, and the smallest *P*-value along with the corresponding haplotype was recorded. Simulation *P*-values were calculated from empirical null distributions based on at least 10,000 permutations. A haplotype frequency cutoff of 0.005 was used for all analyses.

To adjust *P*-values produced by Fisher's exact test for multiple testing, we used the correction of Benjamini and Hochberg [1995]. To adjust the *P*-values produced in haplotype analyses, we used a permutation procedure to simulate *P*-values

[Zhao et al., 2000]. Disease status (cases vs. controls) was randomly permuted among the persons tested, keeping their SNP genotypes unaltered; the lowest simulation *P*-value was calculated for each of 10,000 permutations for the data set for several selected window sizes, and from this distribution we estimated the *P*-value that corresponded to the conventional 5% threshold. All SNPs were subjected to a Hardy-Weinberg equilibrium test in the control group, using Fisher's exact test.

RESULTS

MAPPING IN THE "VALIDATION" SET: LINKAGE DISEQUILIBRIUM IN THE *BRCA2* REGION

In order to evaluate the strength and extent of LD in the region of *BRCA2* in carriers of the *BRCA2∗6174delT* mutation, the 8,576 SNPs were analyzed in 27 unrelated mutation carriers and 63 healthy Ashkenazi Jews as controls. Performing Fisher's exact test of the three genotypes at each individual SNP locus, we observed two loci (TSC1378449 and TSC599767) on chromosome 13 in the region that contains *BRCA2* that displayed unadjusted *P*-values of 1.7×10^{-9} and 5.0×10^{-5}, respectively (Fig. 1A). Because multiple tests were performed, we adjusted the *P*-values with the correction of Benjamini and Hochberg [1995] (Fig. 1B). The *P*-values obtained at both TSC1378449 and TSC599767 were significant after adjustment (3.5×10^{-7} and 1.0×10^{-2}, respectively), and only one other *P*-value obtained in the analysis was smaller than the conventional $P < 0.05$ cutoff (TSC1136736 on chromosome 2 had a *P*-value of 0.05). The SNPs TSC1378449 and TSC599767 (2.78 Mbs apart) were particularly informative due to the fact that the *BRCA2∗6174delT* mutation was associated with alleles that were present at low frequencies in controls (0.09 and 0.07, respectively). We noted that TSC1378449 and TSC599767 are separated by 17 SNPs, and a significant adjusted *P*-value was not obtained at any of the 17 intervening SNPs.

We next estimated haplotype frequencies for a sliding window of n loci (n = 2–12) using the EM algorithm, and calculated *P*-values associated with each haplotype by the maximum likelihood ratio test (Fig. 1C). Minimum *P*-values less than 1×10^{-8} were obtained in the *BRCA2* region for 10 of 11 window sizes. Other than the *P*-values in the *BRCA2* region, no other *P*-values in the analysis were smaller than 1×10^{-5}. Taking into account multiple testing, we estimated by permutation analysis, using a window size of 8 consecutive loci, that a *P*-value less than 6×10^{-6} would fall below the 5th percentile of *P*-values. (We noted that this correction was similar to using the Bonferroni correction.) Consequently, in the known mutation carriers, the *P*-values obtained in the region containing *BRCA2* were highly significant, and no other *P*-values in the analysis achieved significance.

In haplotype analyses, the smallest *P*-values were obtained for haplotypes that included the TSC1378449 and TSC599767 loci (Fig. 2), indicating that the power of haplotype analyses was also driven by the allele frequencies of the loci tested. In the haplotype analysis, a series of overlapping segments in LD with *BRCA2* was identified spanning a 10-Mb region and containing 48 SNPs (Fig. 2). Most chromosomes (25/27) shared alleles in a 1-Mb region flanking *BCRA2*. On the proximal side of *BRCA2*, breaks in the founder haplotype were identified on 8 chromosomes in the first Mb proximal to *BRCA2*, 9 in the second, 7 in the third, and 2 in the fourth. On the distal side, breaks were identified on 6 chromosomes in the first Mb distal to *BRCA2*, 1 in the second, 3 in the third, 9 in the fourth, and 8 in the fifth. The average extent of the shared region on *BRCA2∗6174delT* chromosomes was 5.6 Mbs. The broad extent of allele-sharing in the *BRCA2* region reflects the relatively recent origin of the *BRCA2∗6174delT* chromosome in the Jewish population, and the multiplicity of breaks in the founder haplotype represents historical recombination events that flank the *BRCA2∗6174delT* mutation. The proximity of breaks flanking *BRCA2* could be used to localize the gene to an approximately 1-Mb critical region.

MAPPING IN THE "FIELD" SET DIMINISHED POWER DUE TO HETEROGENEITY

In order to evaluate the power of genome-wide SNP LD mapping in detecting associations in the *BRCA2* region, 8,576 SNPs were analyzed in 19 Jewish breast cancer cases, drawn from breast cancer kindreds defined by the presence of a male breast cancer, and in the aforementioned 63 healthy Jewish controls. In this ascertainment, 12 of 19 cases carried the *BRCA2∗6174delT* mutation. Performing Fisher's exact test at each individual SNP locus, we observed a *P*-value of 4.5×10^{-4} at TSC1378449, and of 1 at TSC599767 (Fig. 3A). This *P*-value was considerably less than the value obtained ($P = 3 \times 10^{-7}$) when the 12 cases who

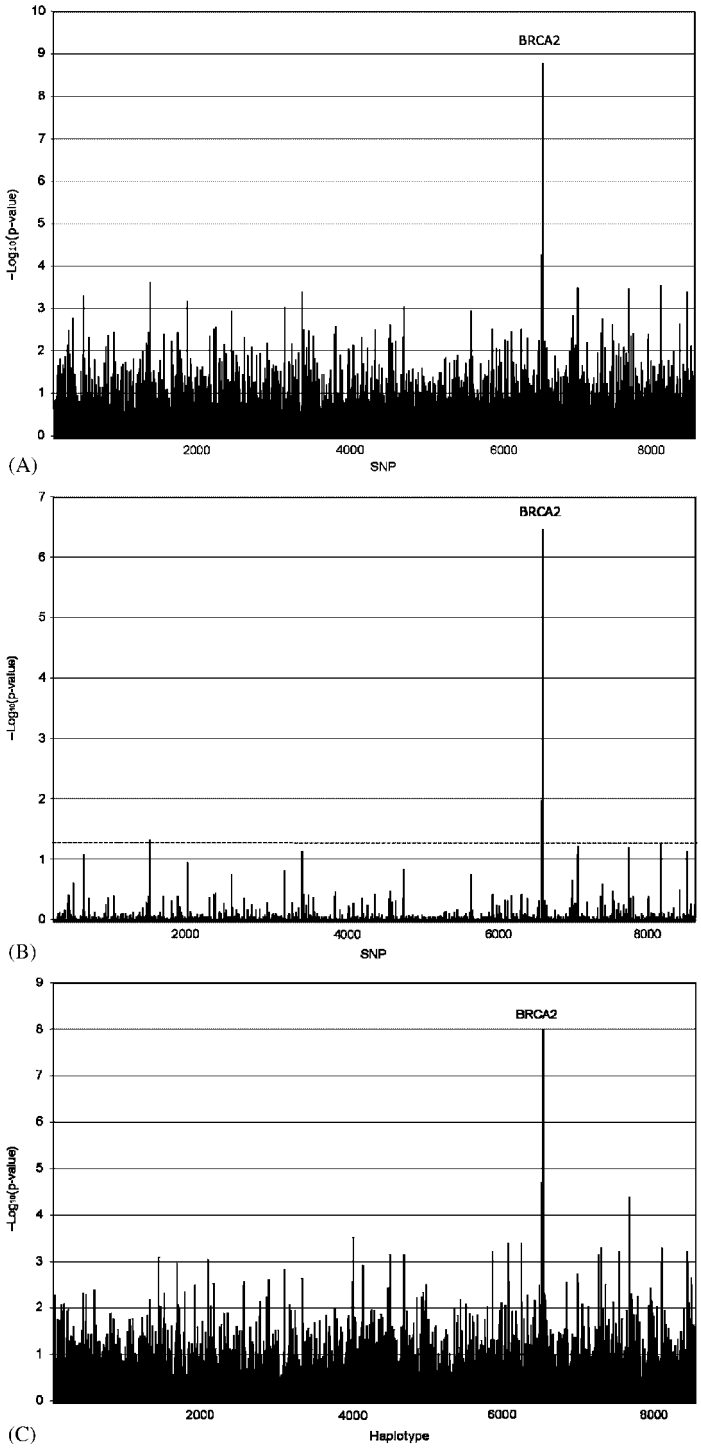

Fig. 1. A: Plot of *P*-values obtained by two-sided Fisher's exact test comparing genotype frequencies (2×3) at each of the 8,576 SNP loci tested in the "validation" set: 27 Ashkenazi Jewish *BRCA2*6174delT* carriers with breast cancer vs. 63 healthy Ashkenazi Jewish controls. Along x-axis, loci were arranged by nucleotide number, starting at nucleotide 1 on chromosome 1, and ascending by chromosome number to the last nucleotide on chromosome 22. *P*-values were transformed by $-\log_{10}$(*P*-value) to display scores as peaks. The smallest *P*-value in the analysis ($P = 1.7 \times 10^{-9}$) was obtained at TSC1378449 at nucleotide position 31,468,145 on chromosome 13 (human genome database build 33). Region containing *BRCA2* is identified. B: Correction of Benjamini and Hochberg [1995] of *P*-values obtained by Fisher's exact test shown in A. Dashed line represents conventional 0.05 *P*-value cutoff. C: Plot representing minimum *P*-values calculated by maximum likelihood ratio test after estimating frequencies of 8-locus haplotypes by EM algorithm in cases vs. controls. The smallest *P*-value ($P < 10^{-8}$) was obtained in *BRCA2* region at three consecutive haplotypes TSC208530–TSC1378449, TSC1083028–TSC64502, and TSC195208–TSC983347, all three of which contain locus TSC1378449 identified by Fisher's exact test.

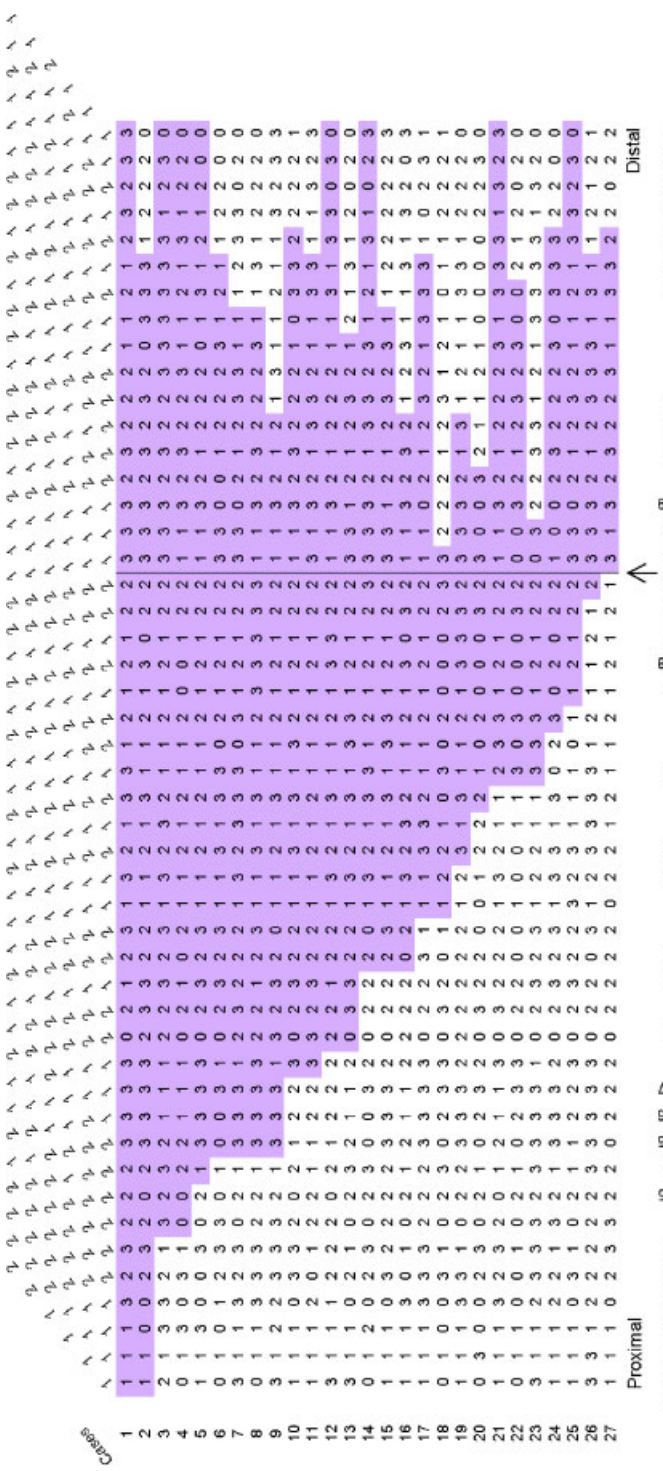

Fig. 2. LD structure of chromosomes that carry *BRCA2*6174delT* mutation. Shown across top are 6-locus haplotypes that had minimum *P*-values by maximum likelihood ratio test for 48 consecutive loci that are contained in a region of LD surrounding *BRCA2*. For representation of haplotypes, alleles at each locus were arbitrarily assigned a number 1 or 2. Below the array of haplotypes are shown genotypes at each of 48 loci for each of 27 cases (1, homozygous for allele 1; 2, homozygous for allele 2; 3, heterozygous; 0, missing data). Portion of founder haplotype that is carried by each case is shaded: a break was identified by the locus nearest to *BRCA2* at which a nonfounder allele was present on the chromosome; all loci distal to the break relative to *BRCA2* were not shaded. Arrow indicates position of *BRCA2* gene. Nucleotide positions are given from human genome database build 33. Row Hap p contains *P*-values for each haplotype shown. Leftmost *P*-value shown, which is represented by transformation $-\log_{10}(P\text{-value})$, was obtained for leftmost haplotype shown containing SNPs TSC987224–TSC128714, next *P*-value to right for next haplotype to right containing SNPs TSC273115–TSC56508, and so forth. At some loci, more than one allele appeared associated with founder haplotype, which indicated that ambiguity in the founder allele exists at these loci. Row SNP p contains *P*-values obtained for each SNP by Fisher's exact test. Rows F(case) and F(control) contain frequencies of minor allele of each SNP in cases and controls, respectively.

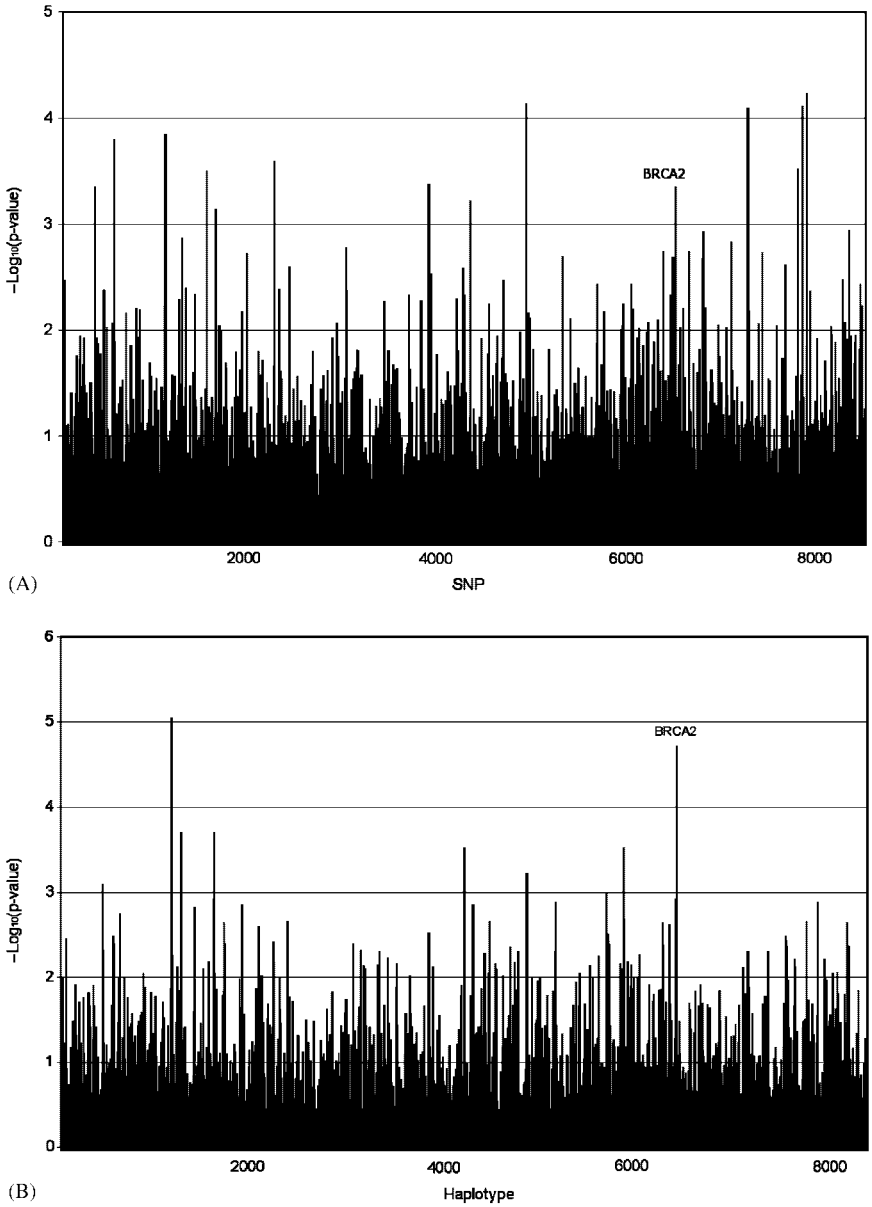

Fig. 3. A: Plot of *P*-values obtained by two-sided Fisher's exact test comparing genotype frequencies (2×3) at each of 8,576 SNP loci tested in "field" set: 19 Ashkenazi Jewish breast cancer cases from male breast cancer families, in comparison to 63 healthy Ashkenazi Jewish controls. The smallest *P*-value obtained in analysis (5.9×10^{-5}) was at TSC57977 on chromosome 18. TSC1378449 in the *BRCA2* region obtained *P*-value of 4.5×10^{-4}, which was the 12th smallest in analysis. B: Plot representing minimum *P*-values calculated by maximum likelihood ratio test after estimating frequencies of 8-locus haplotypes by EM algorithm in 19 cases vs. 63 controls. The smallest *P*-value (8.9×10^{-6}) obtained in analysis was at 8-locus haplotype TSC865709–TSC261312 on chromosome 2. The second smallest *P*-value (1.9×10^{-5}) was obtained at haplotype TSC991493–TSC43705, which contains SNP TSC1378449 and the *BRCA2* locus. The adjacent haplotype proximal to this one TSC815459–TSC905266 recorded 3rd smallest *P*-value (1.5×10^{-4}) in the 8-locus haplotype analysis.

carried the *BRCA2∗6174delT* mutation were compared to the 63 controls. The *P*-value at TSC1378449 was the 12th smallest of the 8,576 *P*-values determined. The smallest *P*-value obtained was 5.9×10^{-5} at TSC57977 on chromosome 18. None of the *P*-values obtained in this analysis were significant after the correction of Benjamini and Hochberg [1995] was applied (data not shown).

In the haplotype analysis, the smallest *P*-values in the *BRCA2* region were associated with the haplotypes that contained the SNP TSC1378449

(Fig. 3B). The *BRCA2* region had the smallest *P*-value in the analysis with window sizes of 3 and 4, and one of the three smallest *P*-values with window sizes of 3–9. In absolute terms, the minimum *P*-value for the region containing TSC1378449 ranged from 0.002 (window size 12) to 1.9×10^{-5} (window size 8). None of the *P*-values obtained in the *BRCA2* region were considered significant after adjustment for multiple testing. Only one locus, TSC261313 on chromosome 2, had a significant *P*-value (5.5×10^{-6}), which was observed in the analysis of seven-locus haplotypes; presumably, it is a false-positive result. Only two regions had *P*-values smaller than the smallest *P*-value obtained for the *BRCA2* region. In summary, these results showed that introduction of even a modest percentage of genetic heterogeneity resulted in a large reduction in power to detect significant associations. However, because the tests for association of haplotypes in the *BRCA2* region were among the most significant in the analysis across varying haplotype sizes, the results suggest that a strategy to identify new loci would be to focus on regions in which small *P*-values were obtained in multiple window sizes.

GENOME-WIDE SNP LD MAPPING OF THE "TEST" SET APPROACH TO NOVEL SUSCEPTIBILITY LOCI

To detect novel susceptibility loci, we analyzed 8,576 SNPs in 57 breast cancer cases drawn from kindreds, in which four or more persons (inclusive of the proband) had been diagnosed with breast cancer and in whom mutations in *BRCA1* and *BRCA2* were not present, and in the 63 healthy controls. We then compiled a table of all loci that obtained *P*-values of 10^{-3} or smaller; 10^{-3} was chosen as an arbitrary cut-point.

In the single-marker analysis, a total of 8 loci obtained *P*-values of 10^{-3} or smaller, of which 2 were also identified in haplotype analyses. The smallest *P*-value obtained was 3.8×10^{-4} at TSC51781 on chromosome 17 (Fig. 4A). None of the *P*-values were significant after the correction of Benjamini and Hochberg [1995] (data not shown). A *P*-value of 6×10^{-4} was obtained at the SNP TSC671100 at 26.8 Mbs on chromosome 8p12, which is within a region previously identified by linkage analysis [Kerangueven et al., 1995; Seitz et al., 1997].

In haplotype analyses, a total of 92 haplotypes obtained *P*-values of 10^{-3} or smaller. When we took into account the overlapping haplotypes and nonoverlapping haplotypes within 2 Mbs of each other, 39 different regions were identified, but only 19 regions were identified with small *P*-values in two or more haplotype windows. The two smallest *P*-values in the entire analysis (both 7.0×10^{-6}) were obtained at the 9-locus haplotype TSC582885–TSC948257 on chromosome 15, and at the 10-locus haplotype TSC181069–TSC1555460 on chromosome 9 (Fig. 4B). The chromosome 15 region had small *P*-values at 13 overlapping haplotypes, the associated haplotypes spanning a 3.9-Mb region from nucleotides 89.0–92.9 Mbs. This region also recorded the third (1.5×10^{-5}) and fourth (2.0×10^{-5}) smallest *P*-values in the analysis, which were the smallest *P*-values obtained in the 11- and 7-locus haplotype analyses, respectively. The chromosome 9 region had small *P*-values at 6 overlapping haplotypes, spanning a 5.6-Mb region from nucleotides 68.2–73.8 Mbs. These *P*-values were just short of the significance threshold of 6.0×10^{-6} set by the permutation analysis. None of the individual SNPs in the chromosome 15 region or in the chromosome 9 region obtained a *P*-value less than 10^{-3} in the single-marker analysis; however, the chromosome 15 region contained six SNPs close together, with *P*-values between 0.09 and 0.01 (Table III). None of the SNPs in the chromosome 9 region obtained a *P*-value less than 0.05. None of the regions identified in the single-marker or haplotype analyses with *P*-values less than 10^{-3} contained loci that failed the Hardy-Weinberg equilibrium test (data not shown).

In our previous analysis of the "field" set, we determined that 4 of the 57 high-risk cases carried the *CHK2* S438F mutation [Shaag et al., 2005]. Although a 9-locus haplotype TSC57163–TSC242417 on chromosome 22 that is less than 5 Mbs from the *CHK2* locus obtained a *P*-value of 6×10^{-4} (data not shown), the associated haplotype was not carried by any of the S438F carriers. An analysis of the haplotypes that contained *CHK2* in the four S438F carriers revealed the presence of a shared haplotype that was common among controls.

DISCUSSION

Detecting significant associations in a common disease such as breast cancer poses significant challenges to the association design. We demon-

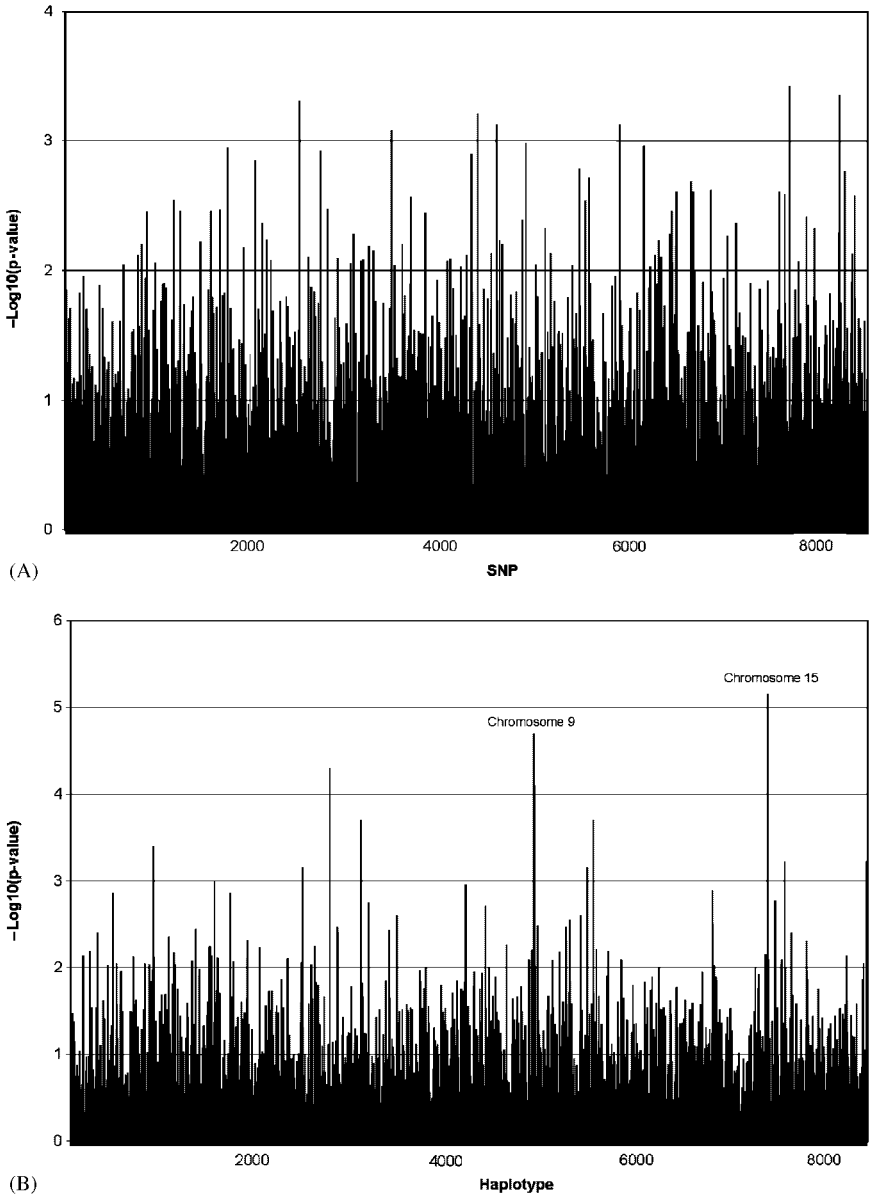

Fig. 4. A: Plot of *P*-values obtained by two-sided Fisher's exact test comparing genotype frequencies (2×3) at each of 8,487 SNP loci tested in the "test" set: 57 Ashkenazi Jewish breast cancer cases from families with 4 or more cases of breast cancer, in whom mutation in *BRCA1* and *BRCA2* was not present, vs. 63 healthy Ashkenazi Jewish controls. The smallest *P*-value (3.8×10^{-4}) was obtained at TSC51781 on chromosome 17. No *P*-values were significant after correction for multiple testing. B: Plot representing minimum *P*-values calculated by the maximum likelihood ratio test after estimating frequencies of 9-locus haplotypes by EM algorithm in the "test" set. The two smallest *P*-values in entire analysis (both 6×10^{-6}) were obtained at the haplotype TSC582885–TSC948257 on chromosome 15 (peak labeled), and at the 10-locus haplotype TSC181069–TSC1555460 on chromosome 9. Second and fourth smallest *P*-values obtained in the analysis of 9-locus haplotypes were obtained at two abutting haplotypes TSC1761488–TSC55652 ($P = 2.5 \times 10^{-5}$) and TSC181009-TSC601577 ($P = 7.9 \times 10^{-5}$) on chromosome 9 (peak labeled).

strated here that genome-wide SNP LD mapping in high-risk cases from a genetically isolated population should be useful for the identification of novel genes, because this study design ensures that the disease-causing mutation and its surrounding genomic environs are identical by descent from a common founder. The data analysis indicated that under certain conditions, genome-wide SNP LD mapping could be successful in the identification of new genes. However, our results also strongly emphasized the pitfalls of the approach, including lower power due to genetic heterogeneity and the adjustments required for multiple testing, and problems arising

TABLE III. *P*-values obtained by two-sided Fisher's exact test in chromosome 15 region[a]

TSC identification	Nucleotide	*P*-value
42322	88988579	0.46
1216920	89002256	0.85
582885	89017827	0.092
582887	89047985	0.043
545438	89176825	0.082
545439	89177135	0.043
1587048	89669783	0.67
52189	90202548	0.92
548898	90312845	0.012
55180	91110486	0.84
948257	91111128	0.56
934302	91506180	0.31
57131	92243342	0.23
57054	92878211	0.031

[a]TSC, the SNP Consortium.

from the frequency of disease alleles in the case group, which is related to their penetrance, and the frequency of the haplotypes on which the disease alleles arose. We discuss each of these limitations at greater length below.

SNP SELECTION

The markers in the Affymetrix panel were selected from SNPs previously validated by the SNP Consortium and for their genome-wide coverage [Kennedy et al., 2003; Matsuzaki et al., 2004]. Each SNP had a relatively high heterozygosity (>0.25) in three population groups, and the SNPs were selected without consideration for proximity to genes. However, the difference between the mean and median distances between SNPs indicates that the SNPs were not evenly spaced. Because the density of SNPs in this panel was low and uneven, false-negative results could be obtained if an appropriate SNP or group of SNPs was not present in the panel. Despite these problems, in a total of three similarly designed proof-of-principle experiments performed by our group (one on Jewish Bloom's syndrome, one on Jewish HNPCC [Mitra et al., 2004], and the present one on Jewish hereditary breast cancer syndrome), we were able to detect strong associations between single markers or haplotypes and the disease genes responsible for these syndromes using a low-density SNP panel.

The reason we were able to detect associations in our "validation" set was derived from two features of the design: 1) a high proportion of cases carried the disease allele, and 2) specific low-frequency SNPs were by chance tested in the large LD blocks that encompassed the disease alleles. Our data showed that the frequency of marker alleles in controls that were associated with the disease-causing mutation was an important factor in obtaining small *P*-values. At two of the SNPs in the *BRCA2* region (TSC1378449 and TSC599767), the *BRCA2∗6174delT* mutation was found in association with the less frequent allele in the controls (minor allele frequencies in controls of 0.09 and 0.07, respectively). In single-marker analysis of *BRCA2∗6174delT* carriers and healthy controls, *P*-values of 1.7×10^{-9} and 5.3×10^{-5} were obtained at TSC1378449 and TSC599767, respectively. TSC1378449 is 822 kb distal to *BRCA2*, and it was present on 25 of 27 *BRCA2∗6174delT* chromosomes examined. The presence of this SNP in the panel was essential for obtaining small *P*-values in the *BRCA2* region in the "field" sets. The *P*-value obtained at TSC1378449 by Fisher's exact test in the subset of 12 *BRCA2∗6174delT* carriers in the "field" set was 1×10^{-7}, which was less than the *P*-value obtained in the analysis of 27 carriers in the "validation" set on account of the reduction in the sample size of cases, but it was much greater than the *P*-value (4.5×10^{-4}) obtained for the analysis of the complete "field" set, demonstrating the effects on power of sample size reduction and genetic heterogeneity. TSC599767 is 1.75 Mb proximal to *BRCA2* and was present on only 13 of the 27 *BRCA2∗6174delT* chromosomes examined, which explains its diminished strength to detect association in the "field" set.

The analysis of the *CHK2* region in the "test" set revealed a pitfall that may be encountered in genome-wide SNP LD mapping in the setting of association studies of genetic isolates. Four cases (7%) in the "test" set carried the S438F allele, compared to the ~1% population frequency in Ashkenazi Jews or the ~2% frequency in consecutive breast cancer cases [Shaag et al., 2005 and unpublished observations]. Although the density of SNPs in the *CHK2* region was close to the mean for the entire collection, the SNPs used did not provide sufficient power to discriminate the S438F allele from the common haplotype on which the mutation arose. A larger sample size of similarly selected cases and additional SNPs in the region would have been needed to identify an association between Ashkenazi Jewish breast cancer and the *CHK2* S438F allele.

Because the possibility of detecting association depends on observing a frequency difference

between cases and controls, a low-frequency allele in complete LD with a disease-causing mutation has the most power for gene discovery. Thus, the presence of low-frequency alleles in the SNP panel could be important for gene discovery, especially if the cancer susceptibility alleles, like *CHK2* S438F, are present at low frequencies. Using SNP panels with greater numbers of SNPs will increase the likelihood that such associated SNPs are present in the panel. It was estimated that over 500,000 SNPs genome-wide will be required to efficiently map low-penetrance disease alleles. However, it is currently not known which are the best SNPs to include in such a panel, and the penalty for using more SNPs is a larger adjustment of *P*-values to compensate for multiple testing. Because testing more SNPs decreases the power of the study design, it is all the more important to select high-risk cases so that the highest proportion of cases possible carry the disease alleles being sought.

P-VALUE CUTOFFS AND REPLICATION

Several weaknesses were inherent in the design of our study. These weaknesses were: 1) cases were taken from two different geographic locales, but matched controls were not included from one of these locales; 2) the controls used were not properly matched to cases by the standard epidemiologic criteria (some controls were male; some were from Israel); 3) the SNP panel did not provide even coverage of the entire genome, and was not informative of all variation in the human genome; and 4) the study was underpowered to detect disease-causing alleles in *BRCA2* or other genes. Our study design had these weaknesses primarily because genome-wide genotyping is expensive, which limited both the number of cases and number of properly matched controls we were able to genotype. Nonetheless, in the analysis of the ''test'' data set, *P*-values close to the significance threshold obtained after correction for multiple testing were achieved in two regions of the genome: one on chromosome 15, and another on chromosome 9. A third region was identified in the haplotype analysis that was identified previously by linkage analysis [Kerangueven et al., 1995; Seitz et al., 1997]. To determine whether these potential associations can be replicated, haplotype-tagging SNPs from the three candidates regions should be genotyped in the ''test'' set, and thereafter the most informative SNPs should be tested in an independently ascertained series of breast cancer cases and controls.

In the haplotype analysis, when multiple testing was taken into account, *P*-values less than 6×10^{-6} were considered statistically significant. Possible false-positive results were suggested by the results of the haplotype analysis of the ''field'' set, in which a *P*-value of 5×10^{-6} was obtained at TSC261313 on chromosome 2. Formally, such results could have been caused by real associations, such as with modifier genes. However, with a similar analytical strategy using samples from Ashkenazi Jewish colon cancer families, we also obtained several significant associations outside of the *MSH2* region [Mitra et al., 2004] that contains a founder mutation in Ashkenazi Jews [Foulkes et al., 2002]. Consequently, even after adjustments are made for multiple testing, regions could be identified with putatively significant *P*-values that are false positives. The best way to counter this problem is to increase the sample size, which would reduce the chances of finding false positives and increase the power of the analysis (the chances of detecting real positives with significant *P*-values).

In comparison to conventional linkage approaches, the association design tested had comparable power for gene localization. With respect to identification of the minimum regions that contain these genes, the association method defined a smaller critical region relative to the numbers of cases studied. With additional SNP genotyping, a minimum region of LD flanking the *BRCA2∗6174delT* mutation of *BRCA2* could be as small as 1 Mb or smaller, whereas an analysis of many meioses would be required in linkage analysis to obtain a similarly small critical region. Moreover, because this SNP LD mapping method required only a single case per kindred, it is easier to collect samples for analysis, making it better suited than linkage for diseases in which family material is difficult to collect (e.g., in late-onset and potentially lethal disorders such as cancer).

While genome-wide SNP LD mapping presents opportunities for the identification of new cancer susceptibility alleles, significant challenges remain regarding the optimal density and distribution of SNPs and the optimal size and type of study design. The results of the current analysis suggest both the pitfalls and promises of this approach to mapping breast cancer susceptibility genes. Based on our findings, we suggest that the keys to success in future studies are 1) selection of a large number of high-risk cases, defined by family

history or multiple cancer phenotype, thus enriching for cancer susceptibility alleles; 2) selection of cases from a genetically isolated population to reduce allelic heterogeneity and to increase LD block size; 3) selection of SNPs that appropriately sample variation across the genome, possibly with an emphasis on rare haplotypes; 4) application of appropriate corrections for multiple testing to control for false positives; and 5) validation of potential associations in the "enriched" case series by testing an independently ascertained series of unselected breast cancer cases. However, even with this two-stage strategy, thousands of unselected cases would be required to confirm associations [Houlston and Peto, 2003]. Such large-scale two-stage studies to localize breast cancer susceptibility loci by SNP LD mapping are underway in the US and UK. Taking each of the above-mentioned five considerations into account in these studies will be necessary to provide the greatest chance of identifying both intermediate- and low-penetrance breast cancer susceptibility alleles.

ACKNOWLEDGMENTS

We are grateful to Agnes Viale and the Genomics Core Facility at Memorial Sloan-Kettering Cancer Center for assistance with genotyping. The authors thank Peter Gregersen and the New York Cancer Project, which is supported by the Academic Medicine Development Company, for providing DNA from healthy Jewish controls. This work was supported by the Lymphoma Foundation's Werner and Elaine Dannheisser Fund, the Frankel Fellowship, the Danziger Foundation, the Byrne Foundation, the Sloan-Kettering Institute, and National Institutes of Health grants R01-HL56778 (N.A.E.), CA103394 (N.M.), and K01-HG00055 and MH44292 (D.G.).

REFERENCES

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 57:289–300.

Botstein D, Risch N. 2003. Discovering genotypes underlying human phenotypes: past successes for Mendelian disease, future approaches for complex disease. Nat Genet [Suppl] 33: 228–237.

Charbonnier F, Raux G, Wang Q, Drouot N, Cordier F, Limacher JM, Saurin JC, Puisieux A, Olschwang S, Frebourg T. 2000. Detection of exon deletions and duplications of the mismatch repair genes in hereditary nonpolyposis colorectal cancer families using multiplex polymerase chain reaction of short fluorescent fragments. Cancer Res 60:2760–2763.

Claus EB, Schildkraut JM, Thompson WD, Risch NJ. 1996. The genetic attributable risk of breast and ovarian cancer. Cancer 77: 2318–2324.

Colditz GA, Willett WC, Hunter DJ, Stampfer MJ, Manson JE, Hennekens CH, Rosner BA. 1993. Family history, age, and risk of breast cancer. Prospective data from the Nurses' Health Study. JAMA 270:338–343.

Couch FJ, Weber BL. 2002. Breast cancer. In: Vogelstein B, Kinzler KW, editors. The genetic basis of human cancer. New York: McGraw-Hill. p 549–581.

Excoffier L, Slatkin M. 1995. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. Mol Biol Evol 12:921–927.

Fallin D, Cohen A, Essioux L, Chumakov I, Blumenfeld M, Cohen D, Schork NJ. 2001. Genetic analysis of case/control data using estimated haplotype frequencies: application to *APOE* locus variation and Alzheimer's disease. Genome Res 11:143–151.

Ford D, Easton DF, Stratton M, Narod S, Goldgar D, Devilee P, Bishop DT, Weber B, Lenoir G, Chang-Claude J, Sobol H, Teare MD, Struewing J, Arason A, Scherneck S, Peto J, Rebbeck TR, Tonin P, Neuhausen S, Barkardottir R, Eyfjord J, Lynch H, Ponder BAJ, Gayther SA, Birch JM, Lindblom A, Stoppa-Lyonnet D, Bignon Y, Borg A, Hamann U, Haites N, Scott RJ, Maugard CM, Vasen H, Seitz S, Cannon-Albright LA, Schofield A, Zelada-Hedman M. 1998. Genetic heterogeneity and penetrance analysis of the *BRCA1* and *BRCA2* genes in breast cancer families. The Breast Cancer Linkage Consortium. Am J Hum Genet 62:676–689.

Foulkes WD, Thiffault I, Gruber SB, Horwitz M, Hamel N, Lee C, Shia J, Markowitz A, Figer A, Friedman E, Farber D, Greenwood CM, Bonner JD, Nafa K, Walsh T, Marcus V, Tomsho L, Gebert J, Macrae FA, Gaff CL, Paillerets BB, Gregersen PK, Weitzel JN, Gordon PH, MacNamara E, King MC, Hampel H, De La Chapelle A, Boyd J, Offit K, Rennert G, Chong G, Ellis NA. 2002. The founder mutation *MSH2*∗*1906G→C* is an important cause of hereditary nonpolyposis colorectal cancer in the Ashkenazi Jewish population. Am J Hum Genet 71:1395–1412.

Frank TS, Deffenbaugh AM, Reid JE, Hulick M, Ward BE, Lingenfelter B, Gumpper KL, Scholl T, Tavtigian SV, Pruss DR, et al. 2002. Clinical characteristics of individuals with germline mutations in BRCA1 and BRCA2: analysis of 10,000 individuals. J Clin Oncol 20:1480–1490.

Gruber SB, Ellis NA, Scott KK, Almog R, Kolachana P, Bonner JD, Kirchhoff T, Tomsho LP, Nafa K, Pierce H, Low M, Satagopan J, Rennert H, Huang H, Greenson JK, Groden J, Rapaport B, Shia J, Johnson S, Gregersen PK, Harris CC, Boyd J. 2002. *BLM* heterozygosity and the risk of colorectal cancer. Science 297:2013.

Houlston RS, Peto J. 2003. The future of association studies of common cancers. Hum Genet 112:434–435.

Houlston RS, Peto J. 2004. The search for low-penetrance cancer susceptibility alleles. Oncogene. 23:6471–6476.

Kainu T, Juo SH, Desper R, Schaffer AA, Gillanders E, Rozenblum E, Freas-Lutz D, Weaver D, Stephan D, Bailey-Wilson J, Kallioniemi OP, Tirkkonen M, Syrjäkoski K, Kuukasjärvi T, Koivisto P, Karhu R, Holli K, Arason A, Johannesdottir G, Bergthorsson JT, Johannsdottir H, Egilsson V, Barkardottir RB, Johannsson O, Haraldsson K, Sandberg T, Holmberg E, Grönberg H, Olsson H, Borg A, Vehmanen P, Eerola H, Heikkilä J, Pyrhönen S, Nevanlinna H. 2000. Somatic deletions in hereditary breast cancers implicate 13q21 as a putative novel breast cancer susceptibility locus. Proc Natl Acad Sci USA 97:9603–9608.

Kauff ND, Perez-Segura P, Robson ME, Scheuer L, Siegel B, Schluger A, Rapaport B, Frank TS, Nafa K, Ellis NA, Parmigiani G, Offit K. 2002. Incidence of non-founder *BRCA1* and *BRCA2* mutations in high risk Ashkenazi breast and ovarian cancer families. J Med Genet 39:611–614.

Kennedy GC, Matsuzaki H, Dong S, Liu WM, Huang J, Liu G, Su X, Cao M, Chen W, Zhang J, Liu W, Yang G, Di X, Ryder T, He Z, Surti U, Phillips MS, Boyce-Jacino MT, Fodor SP, Jones KW. 2003. Large-scale genotyping of complex DNA. Nat Biotechnol 21:1233–1237.

Kerangueven F, Essioux L, Dib A, Noguchi T, Allione F, Geneix J, Longy M, Lidereau R, Eisinger F, Pebusque MJ, Jacqumier J, Bonaitipellie C, Sobol H, Birnbaum D. 1995. Loss of heterozygosity and linkage analysis in breast carcinoma: indication for a putative third susceptibility gene on the short arm of chromosome 8. Oncogene 10:1023–1026.

Matsuzaki H, Loi H, Dong S, Tsai YY, Fang J, Law J, Di X, Liu WM, Yang G, Liu G, Huang J, Kennedy GC, Ryder TB, Marcus GA, Walsh PS, Shriver MD, Puck JM, Jones KW, Mei R. 2004. Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array. Genome Res 14:414–425.

Meijers-Heijboer H, van den Ouweland A, Klijn J. 2002. CHEK2-Breast Cancer Consortium. Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. Nat Genet 31:55–59.

Mitra N, Ye T-Z, Smith A, Chuai S, Kirchhoff T, Peterlongo P, Nafa K, Phillips MS, Offit K, Ellis NA. 2004. Localization of cancer susceptibility genes by genome-wide SNP linkage disequilibrium mapping. Cancer Res 64:8116–8125.

Neuhausen SL, Gilewski T, Norton L, Tran T, McGuire P, Swensen J, Hampel H, Borgen P, Brown K, Skolnick M, Shattuck-Eidens D, Jhanwar S, Goldgar D, Offit K. 1996a. Recurrent *BRCA2 6174delT* mutations in Ashkenazi Jewish women affected by breast cancer. Nat Genet 13:126–128.

Neuhausen SL, Mazoyer S, Friedman L, Stratton M, Offit K, Caligo A, Tomlinson G, Cannon-Albright L, Bishop T, Kelsell D, Solomon E, Weber B, Couch F, Struewing J, Tonin P, Durocher F, Narod S, Skolnick MH, Lenoir G, Serova O, Ponder B, Stoppa-Lyonnet D, Easton D, King MC, Goldgar DE. 1996b. Haplotype and phenotype analysis of six recurrent *BRCA1* mutations in 61 families: results of an international study. Am J Hum Genet 58:271–280.

Neuhausen SL, Godwin AK, Gershoni-Baruch R, Schubert E, Garber J, Stoppa-Lyonnet D, Olah E, Csokay B, Serova O, Lalloo F, Osorio A, Stratton M, Offit K, Boyd J, Caligo MA, Scott RJ, Schofield A, Teugels E, Schwab M, Cannon-Albright L, Bishop T, Easton D, Benitez J, King MC, Ponder BA, Weber B, Devilee P, Borg A, Narod SA, Goldgar D. 1998. Haplotype and phenotype analysis of nine recurrent *BRCA2* mutations in 111 families: results of an international study. Am J Hum Genet 62:1381–1388.

Offit K, Gilewski T, McGuire P, Schluger A, Hampel H, Brown K, Swensen J, Neuhausen S, Skolnick M, Norton L, Goldgar D. 1996. Germline *BRCA1* 185delAG mutations in Jewish women with breast cancer. Lancet 347:1643–1645.

Peterlongo P, Nafa K, Lerman GS, Glogowski E, Shia J, Ye T-Z, Markowitz AJ, Guillem JG, Kolachana P, Boyd JA, Offit K, Ellis NA. 2003. *MSH6* germline mutations are rare in colorectal cancer families. Int J Cancer 107:571–579.

Phelan CM, Kwan E, Jack E, Li S, Morgan C, Aube J, Hanna D, Narod SA. 2002. A low frequency of non-founder BRCA1 mutations in Ashkenazi Jewish breast-ovarian cancer families. Hum Mutat 20:352–357.

Rahman N, Teare MD, Seal S, Renard H, Mangion J, Cour C, Thompson D, Shugart Y, Eccles D, Devilee P, Meijers H, Nathanson KL, Neuhausen SL, Weber B, Chang-Claude J, Easton DF, Goldgar D, Stratton MR. 2000. Absence of evidence for a familial breast cancer susceptibility gene at chromosome 8p12–p22. Oncogene 19:4170–4173.

SAS Institute, Inc. 2002. SAS/Genetics® user's guide. Cary, NC: SAS Institute, Inc. p 125.

Schaid D. 2004. Evaluating associations of haplotypes with traits. Genet Epidemiol 27:348–364.

Seitz S, Rohde K, Bender E, Nothnagel A, Kolble K, Schlag PM, Scherneck S. 1997. Strong indication for a breast cancer susceptibility gene on chromosome 8p12–p22: linkage analysis in German breast cancer families. Oncogene 14:741–743.

Shaag A, Walsh T, Renbaum P, Kirchhoff T, Nafa K, Shiovitz S, Mandell JB, Welsch P, Lee MK, Ellis N, Offit K, Levy-Lahad E, King MC. 2005. Functional and genomic approaches reveal an ancient CHEK2 allele associated with breast cancer in the Ashkenazi Jewish population. Hum Mol Genet 14:555–563.

Shifman S, Darvasi A. 2001. The value of isolated populations. Nat Genet 28:309–310.

Slattery ML, Kerber RA. 1993. A comprehensive evaluation of family history and breast cancer risk. The Utah Population Database. JAMA 270:1563–1568.

Stratton MR, Ford D, Neuhasen S, Seal S, Wooster R, Friedman LS, King MC, Egilsson V, Devilee P, McManus R, et al. 1994. Familial male breast cancer is not linked to the BRCA1 locus on chromosome 17q. Nat Genet 7:103–107.

Thompson D, Szabo CI, Mangion J, Oldenburg RA, Odefrey F, Seal S, Barfoot R, Kroeze-Jansema K, Teare D, Rahman N, Daly PA, Smyth E, Ponder BAJ, Peto J, Cannon-Albright L, Easton DF, Goldgar DE. 2002. Evaluation of linkage of breast cancer to the putative BRCA3 locus on chromosome 13q21 in 128 multiple case families from the Breast Cancer Linkage Consortium. Proc Natl Acad Sci USA 99:827–831.

Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, Collins N, Gregory S, Gumbs C, Micklem G. 1996. Identification of the breast cancer susceptibility gene *BRCA2*. Nature 378:789–792.

Zhao JH, Curtis D, Sham PC. 2000. Model-free analysis and permutation tests for allelic associations. Hum Hered 50:133–139.